REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
07-02-2010

2. REPORT TYPE
Final Technical Report

3. DATES COVERED *(From - To)*
May 1st, 2007-November 30th, 2009

4. TITLE AND SUBTITLE
Engineering Awareness

5a. CONTRACT NUMBER
FA9550-07-1-0421

5b. GRANT NUMBER
FA9550-07-1-0421

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER
230391

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Crespi, Valentino
Cybenko, V., George

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Califomia State University Los Angeles, 5151 State University drive, CA 90032-8150
Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755.

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-0023

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Generalized Process Tracking. Defined a rigorous concept of "trackability" of processes in a distributed sensing system. Established fundamental properties of processes and sensing infrastructure that are necessary and sufficient for certain types of trackability to be feasible. Problem addressed and solved: determine the "complexity" of estimating state trajectories of a target process based on a discrete-time sequence of noisy "observations". Conducted a comparative analysis of design methodologies for Agent-Based Systems. Machine Leaming complex processes from data: discovery of a new algorithm to leam Hidden Markov Models (HMMs) from typical realizations of the associated stochastic process. The new method is based on the non-negative matrix factorization (NMF) of higher order Markovian statistics and is structurally different from the classical Baum-Welsh and associated approaches. Cognitive Complexification: development of new methods to shape network communications for preventing covert transmissions from hiding behind the statistics of ordinary traffic.

15. SUBJECT TERMS
Trackability, target tracking, sensor systems, distributed control, multi-agent systems, machine leaming, stochastic processes, hidden Markov models, probabilistic automata, covert channels and covert communications.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	SAR		Valentino Crespi
					19b. TELEPHONE NUMBER *(Include area code)*

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18
Adobe Professional 7.0

Subject: Extended Final Report

Contract/Grant Title: Engineering Awareness **PI:** Valentino Crespi

Contract/Grant #: FA9550-07-1-0421

Reporting Period: 5/1/07 to 11/30/09

Accomplishments: Let us first recall that *Engineering Awareness* is, in our vision, the ability to engineer systems in which effective situational awareness is possible. A fundamental challenge in this art is to be able to effectively monitor environments. Examples of environments include networked computer systems, autonomic computing systems and distributed and dynamic information systems. In our approach an environment consists, in its most abstract form, of multiple processes or behaviors that we typically model as Finite State Machines such as Probabilistic and nonprobabilistic Finite State Automata (DFAs/PFAs), Probabilistic Deterministic Finite State Automata (PDFAs), Probabilistic Suffix Automata (PSAs), Hidden Markov Models (HMMs), etc. The main goal of the project has been to establish fundamental scientific and engineering results to meet the challenge.

Given this premise the description of our main contributions can be articulated along the following four points:

1. **Process Trackability:** we introduced a rigorous notion of *trackability* of processes/ behaviors with sensor networks.

   We developed a quantitative theory of *trackability* of *weak models* that investigates the rate of growth of the number of consistent tracks given a temporal sequence of observations made by the sensor network.

   The phenomenon being tracked is modeled by a nondeterministic finite automaton (a weak model) and the sensor network is modeled by an observer capable of detecting events related, typically ambiguously, to the states of the underlying automaton. Formally, an input string of symbols (the sensor network observations) that is presented to a nondeterministic finite automaton, M, (the weak model) determines a set of state sequences (the tracks or hypotheses) that are capable of generating the input string. We study the growth of the size of this candidate set of tracks as a function of the length of the input string.

   One key result is that for a given automaton and sensor coverage, the worst-case rate of growth is either polynomial or exponential in the number of observations, indicating a kind of phase transition in tracking accuracy. Moreover this character can be decided in polynomial time in the size of the representation of the model. Technically we related this problem to deciding whether the Joint Spectral Radius of a finite set of matrices with entries in $\{0, 1\}$ is less than or equal to 1.

   These results have applications to various tracking problems of recent interest involving tracking phenomena using noisy observations of hidden states such as: sensor networks, computer network security, autonomic computing and dynamic social network analysis.

Those results appeared in a single comprehensive seminal paper [1] which was published after four years of reviews and nontrivial enhancements in the special format of 42 pages (as 30 was the journal limit).

2. **Process Learning:** we devised a novel methodology to machine learn Hidden Markov Models (HMMs) from observed (typical) data. Our new algorithms are based on the *non-negative matrix factorization (NMF)* of higher order Markovian statistics and are structurally different from the classical Baum-Welsh and associated approaches.

   At a conceptual level, our algorithm operates as follows. We first estimate the matrix of an observation sequence's high order statistics. This matrix has a natural non-negative matrix factorization (NMF) which can be interpreted in terms of the probability distribution of future observations given the current state of the underlying Markov Chain. Once estimated, these probability distributions can be used to directly estimate the transition probabilities of the HMM.

   Part of these results are contained in [3]. This work has been accepted for publication and is to appear in IEEE Transactions on Information Theory. The original (first) submission is publicly available on line in ArXiv. The final version is about to be uploaded to ArXiv as well.

   More papers containing the most recent results are in course of preparation, including a monograph on Machine Learning Processes.

3. **Process Complexification:** we developed new methods to shape network communications in order to prevent covert transmissions from hiding behind the statistics of ordinary traffic.

   The general idea is the following. In a local area network we view traffic as a stationary stochastic process. We then machine learn its stationary k-order statistics and build a model of a process that shares the same k-order stationary distributions but possesses different $(k+1)$-order stationary distributions. The constructed model can then be used to shape local transmissions.

   We have achieved nontrivial results in this direction and successfully presented their details during the program reviews. This is still work in progress and those results will be soon submitted for publication.

4. **Design Methodologies and Distributed Sampling:**

   The results of this work were published in [2] and were the culmination of a long standing scientific dispute between two different approaches: the Galstyan and Lerman bottom-up methodology, based on Statistical Physics, and the Cybenko and Crespi top-down methodology, evolved from classical control theory.

   Traditionally, two alternative design approaches have been available to engineers: top-down and bottom-up. In the top-down approach, the design process starts with specifying the global system state and assuming that each component has global knowledge of the system, as in a centralized approach. The solution is then decentralized by replacing global knowledge with communication. In the bottom-up approach, on the other hand, the design starts with specifying requirements and capabilities of individual components, and the global behavior is said to emerge out of interactions among constituent components and between components and the environment.

We performed a comparative study of two design methodologies and showed that under certain assumptions on the communication and the external environment, both bottom–up and top–down methodologies produce very similar solutions.

We demonstrated those ideas on a scenario of distributed sampling: in a closed arena a known number of $M_0 = R + G$ (G are green and R are red) pucks have been disseminated in unknown positions. The numbers of either type of puck, R and G, are unknown and can even change in time. We deploy N robots equipped with a red lamp and a green lamp to sample the pucks. Each robot can be foraging for one type of puck at any given time and its foraging state will be displayed by lamp color to other robots. We assume that robots have a memory buffer of a certain length where they can store their recent observations of pucks and other robots. The goal of the application is to have, on average, the same proportion of red to green robots as the proportion of red to green pucks in the arena. The task is to define color–selection rules based on robots' memory and interaction with other robots and/or environment.

Archival Publications:

1. V.Crespi, G.Cybenko, G.Jiang, "The Theory of Trackability with Applications to Sensor Networks", ACM Transactions on Sensor Networks (special publication: 42 pages, journal limit, 30), May 2008.

2. V. Crespi, A. Galstyan, K. Lerman, "Top-Down vs Bottom-up Methodologies in Multi-Agent System Design", Journal of Autonomous Robots, 2008.

3. G. Cybenko, V. Crespi, "Learning Hidden Markov Models using non-negative matrix factorization". To appear in IEEE Transactions on Information Theory, 2011. First version, submitted in September 2008, is available to the public at arXiv:0809.4086v1, 2008.

   Note: Referenced by other scientists, e.g. http://www.cscs.umich.edu/~crshalizi/notabene/inference-markov.html, as one of the most relevant papers in the area.

4. In preparation:

   - V. Crespi, G. Cybenko, "Statistical Learning of Stochastic Behaviors and Processes". Monograph, currently in preparation.
   - V.Crespi, G. Cybenko, A. Giani, "Cognitive Complexification". Currently in preparation.